Data management

Data Management

Data Management tasks may be performed through the Provider User Management Program (PUMP) or the ECHO DataManagementService, AccessControlService, and Group2ManagementService APIs if the provider would like to write their own tool. For more information regarding PUMP, refer to the Data Partners > Managing Data section of the ECHO website (https://earthdata.nasa.gov/echo/). For more information regarding how to use the ACL and Group Management functionality, refer to the "Group Management & ACLs How To" guide (ECHO_Guide_003), the "ECHO ACLs and Roles" (ECHO_OpsCon_013), and the "ECHO Group Management" (ECHO_OpsCon_014) operations concepts available on the ECHO website (https://earthdata.nasa.gov/library/echo-opscon-documents) .

Access Control Concepts

The ECHO API facilitates allows Data Partners to control access to two types of data objects within the ECHO system:

- **ProviderObjects** – Provider information items which are only managed by members of the Data Partner's team and ECHO Operations. For example, provider orders, order policies, and provider groups.
- **Catalog Items** – Provider metadata items which are made available for discovery and ordering to end users. This includes the control of collection and granule items, but not explicitly browse.

While interacting with the ECHO access control capability, you will need to understand the following basic concepts. Each item is discussed in greater detail in subsequent sections of this document.

- **Groups** – Provider-defined groups of ECHO registered users that may have specific access to Provider Objects or Catalog Items based upon the group's permissions. ECHO Operations also manages system level groups for their own access. Virtual system groups (e.g. guest users) are also available for specific purposes.
- **ACL** – An **A**ccess **C**ontrol **L**ist is associated with each Provider Object or provider-defined set of Catalog Items and lists the permissions granted to a provider or system group.
- **Permissions–** Each Provider Object or Catalog Item has a specific set of permissions (e.g. Create, Read, Update, Delete) that may be granted to a group. Due to the nature of object types, not all objects will have the same available permissions.
- **Catalog Item Identifiers–** Each *CatalogItem* ACL has a unique set of identifiers which are used to determine which catalog items will be affected by the ACL.
- **Provider "Administrators" Group** – Each ECHO Data Partner will have a provider group named "Administrators" which will initially be granted all permissions on all objects and be given the ability to pass on any of their permissions.

- **Restriction Flag** – Metadata values specifying a provider specific restriction value scheme.

Groups

An ECHO Group is identified by the following information:

- **Name** – Unique group name within the Data Partner's list of groups
- **Description** – Description of the purpose of the group and the users which it contains.
- **Members** – List of ECHO registered users who are a part of the group.
- **Provider** – The provider who owns the group, or possibly the system if it is a system level group.

ECHO groups are associated with an ECHO provider and appear only in the context of that provider. The ECHO API supports retrieving groups for a single provider at a time. Group names are scoped to the owning provider, therefore names only need to be unique within the list of groups owned by the provider, and not across all groups in the system. ECHO Data Partners may create as many groups as are needed in order to fulfill their data access control needs.

ECHO also supports the concept of system level groups which are managed by the ECHO Operations team. These groups allow the ECHO Operations team to manage their own access to Provider Objects and Catalog Items without needing to coordinate with each provider. Although ECHO Operations will have all permissions on all objects, they will continue to communicate changes with ECHO Data Partners to ensure good coordination.

In addition to ECHO Operations' managed system groups, ECHO also has the concept of *virtual* system groups. These groups are "managed" by the ECHO system and include a *Registered Users* group and *Guest Users* group. Due to the nature of these lists, ECHO will dynamically associate a user with one of these virtual groups when assessing permissions. Permissions to **both** of these groups must be independently managed.

The permissions to create new groups and view existing groups are managed by assigning permissions on the *Group* Provider Object ACL. Permissions to update or delete existing groups are managed by assigning permissions on the *Group Management* Provider Object ACL. When creating a new group, an initial group must be specified as the *Initial Management Group*. This group will be given permissions to update and delete the new group, and others may be added later.

Access Control List (ACL)

An Access Control List is responsible for linking groups to a specific *Provider Object* or *Catalog Item* and describing the permissions the group has been granted. As was described previously, *Provider Objects* include such things as provider orders, order policies, and groups. *Catalog Item* ACLs are assigned to a custom static or dynamic listing of collection and granule items. All provider objects and catalog items are **not accessible** by default. An ECHO Data Partner uses ACLs to grant permissions to ECHO groups so that group members will have access to specific objects or sets of catalog items. An ACL may exist without any assigned permissions. For example, an ECHO Data Partner may wish to not assign any permissions to the "Extended Services" *Provider Object*, or they may define a *Catalog Item* ACL which controls access to a specific data set, but choose to not assign any permissions at the present time.

Permissions

The permissions for each *Catalog Item* may be *View* or *Order,* without exception. The permissions for each *Provider Object* may be *Create, Read, Update,* or *Delete,* however the available permissions differs depending on the nature of that object. For example, the *Provider Policies* object can be granted the *Read* or *Update* permissions, while the *Provider Audit Report* object can only be granted the *Read* permission. Some permissions, such as the ability to read option definitions, are not grantable on some *Provider Objects* because access is open to any user. The following table outlines the grantable permissions for each *Provider Object*.

Provider Object Grantable Permissions

Provider Object	Grantable Permissions	Description
Audit Report	Read	Allows the viewing of an audit report for actions associated with a specific provider
Dataset Information	Read	Allows the usage of the reconciliation GetDatasetInformation() method.
Extended Services (all types)	Create, Update, Delete	Allows the creation, updating, and deletion of extended services.
Ingest Operations	Read, Update	Controls access to who can log into the EIAT (Not used by ECHO Ingest).
Groups	Create, Read	Allows the creation of new groups or viewing of existing provider groups
Group Management	Update, Delete	Allows the updating or deletion of an existing provider group.

Option Assignments	Create, Read, Delete	Allows the assignment of an option definition to one or more datasets.
Option Definitions	Create, Delete	Allows the creation and deleting of an option definition.
Option Definition Deprecation	Create	Allows the deprecation of an option definition.
Provider Context	Read	Allows a user to act as a provider and perform all permitted provider actions.
Provider Holdings	Read	Allows the viewing of a provider‘s holdings (dataset & granule count).
Provider Information	Update	Allows provider information to be updated.
Provider Orders	Read	Allows the viewing of all orders associated with a specific provider.
Provider Order Resubmission	Create	Allows the resubmission of a provider‘s order
Provider Order Acceptance	Create	Allows the acceptance of a provider‘s order. Order Fulfillment Service Users (EWOC) will use this ACL.
Provider Order Rejection	Create	Allows the rejection of a provider‘s order. Order Fulfillment Service Users (EWOC) will use this ACL.
Provider Order Closure	Create	Allows the closure of a provider‘s order. Order Fulfillment Service Users (EWOC) will use this ACL.
Provider Order Tracking Id	Update	Allows an order to be updated with a provider tracking ID. Order Fulfillment Service Users (EWOC) will use this ACL.
Provider Policies	Read, Update, Delete	Allows the editing of provider policies.
User	Read	Allows the viewing, updating, and deletion of an ECHO user.
Authenticator Definition	Create, Delete	Allows the creation, deletion of provider authenticators. (Not currently being used)

CatalogItemIdentifiers

Each *Catalog Item* ACL will have a specific set of identifiers which designate the catalog items to which the ACL will apply. There are three main areas of identifiers: Catalog Item Type, Collection Identifiers, and Metadata Filters. Each of these areas are described in this section.

Catalog Item Type

When creating a *CatalogItem* ACL, you may choose to have it apply to *Collections*, *Granules*, or *Both.* The catalog item type chosen will determine whether the ACL will apply to collection items, granule items, or both types of items. For instance, if *Collections* is chosen then the ACL will be used by ECHO when granting access to collection items. This is useful if a collection doesn't have granules or if the granules will require different permissions. A catalog item type of Granules indicates that an ACL should be used by ECHO when granting access to granule items. If *Both* is chosen, then the ACL will be used by ECHO when granting access to both collection and granule items. This is useful if the collections and granules will have the same permissions.

Collection Identifiers

Each *CatalogItem* ACL has a list of collections which is used to identify collections or granules within collections to which the ACL applies. Collections may be identified in the following ways:

- **SelectedList** – A static list of collections which must be manually managed.
- **AnyCollection** – A dynamic list of all collections in the Data Partner's holdings
- **CollectionPatternMatching** – Patterns may be selected to perform text matching on the Data Set ID, Short Name, and/or Version ID.

Metadata Filters

In order to facilitate access control of collections and granules based on metadata fields, a *Catalog Item* ACL may contain collection and

granule metadata filters. The filters are based on temporal fields, restriction flag values, or specific granule UR values. Temporal filters may apply to the acquisition, production, ECHO Insert, or ECHO Last Update field and may be described using an intersection, containment, or disjoint comparator. The full listing of filters is included below:

- **Collection Filters**
  - **Temporal Range**
  - **Rolling Temporal Range**
  - **Restriction Flag**
- **Granule Filters**
  - **Temporal Range**
  - **Rolling Temporal Range**
  - **Restriction Flag**
  - **Granule UR Pattern**

Provider "Administrators" Group

Each ECHO Data Partner will be initially configured with a group named "Administrators." This group will be granted all permissions on all *Provider Object* ACLs and the permissions to manage *Catalog Item* ACLs. This group is initially managed by itself and the system "Administrators" group, which is the ECHO Operations team. The provider "Administrators" group may manage *Provider Object* ACLs and grant permissions to other groups, however the ability to manage and grant permissions may not be given to other provider groups. This is done to provide control over the management of *Provider Object* permissions. The ability to grant management of *Catalog Item* ACLs can be granted to other groups. This is distinctly different than *Provider Object* ACL management and is designed to allow for groups such as a Data Partner's User Services team to manage *Catalog Item* ACLs without coordinating with the provider "Administrators."

RestrictionFlag

The restriction flag is a decimal value which is specified in a Data Partner's collection or granule metadata. Data Partners may configure *a Catalog Item* ACL which utilizes the restriction flag value within a record's metadata. The availability of data can then be changed during ECHO Ingest through the usage of partial updates or full metadata replacement. A granule or collection's restriction flag cannot be updated through the API.

Access Control Recommendations

The following recommendations outline some suggested access control mechanisms to facilitate a Data Partner's data management needs, along with those of the ECHO Operations team.

Data Mgmt & User Services Groups

As has been described, a provider "Administrators" group will be created and granted all permissions on all *ProviderObjects*. It is suggested that membership in this group be limited to those individuals who have need of managing access to the provider. There are two distinct "roles" which may be facilitated by provider groups and those are a "Data Management" group and "User Services" group. The "Data Management" group would be given a subset of *ProviderObject* ACL permissions relevant to their job role. The "User Services" group would have a slightly different subset of *ProviderObject* ACL permissions, but have the ability to manage *CatalogItem* ACLs.

Catalog Item ACLs

When initially configuring *CatalogItem* ACLs, the ECHO Operations team will define two ACLs:

- **AllCollections(NoGranules)** – This ACL dynamically applies to all collections within a Data Partner's holdings, and only affects access to collection metadata. The ECHO Operations team uses this ACL to assign view permissions for the WIST valids process.
- **AllCollectionsandGranules** – This ACL dynamically applies to all granules in all collections within a Data Partner's holdings, and affects access to both collection and granule metadata. This ACL is useful for assigning full permissions to members of the Data Partner team. The ECHO Operations team uses this ACL to assign *view* permissions for the System "Administrators" group.

The following *Catalog Item* ACL conditions are also suggested in order to facilitate general data management:

- **Public Collections and Granules** – This ACL dynamically applies to all granules within a static listing of public collections. As collections are added to the provider's holdings, the provider may add the collection to the listing used by this ACL. *View* permissions to this ACL may be granted to the *Registered Users* and *Guest Users* system level groups to allow for data discovery.
- **Orderable Granules** – This ACL dynamically applies to all granules within a static listing of collections within which the granules are orderable. *Order* permissions to this ACL may be granted to the *Registered Users* and *Guest Users* system level groups to allow for order creation and submission.

Reconciliation

Data Partners may perform data reconciliation via two separate methods, via ECHO Ingest processing or the ECHO Data Management Service API. Both of these methods are described in the following sections.

ECHO Ingest

Through the ECHO 10.0 Ingest Schema, Data Partners may perform two types of metadata reconciliation, *metadata verification* and *inventory verification*. These two verification processes are described below:

- **MetadataVerification** – A full reconciliation of a collection or granule metadata item to include verification of all fields. ECHO will automatically attempt to correct differences within its holdings.
- **InventoryVerification** – A shortened reconciliation mechanism which will allow for the identification of inventory items which are missing from the ECHO holdings or should no longer be held by ECHO. Verified inventory items include collections, granules within a specified collection, or browse records associated with granules within a specified collection.

Both of these methods allow Data Partners to take advantage of Ingest's parallelized data processing and detailed reporting mechanism. For a full explanation of how to utilize these reconciliation capabilities, refer to Section 5.7 of this document.

Data Management Service

This historical method uses the GetDatasetInformation method found on the DataManagementService API. Data Partners use the API to request a subset of metadata to be generated by ECHO for granules within a specific collection filtered by temporal range, online availability, browse availability, and visibility. This method utilizes an optimized internal mechanism to pull the necessary information for each granule matching the request. Due to the large amount of data that may be returned by this method, output is generated in an XML file which is delivered via FTP Push to a specified location.

If temporal ranges are specified, the range type field in the granule must be between the start and stop times. If the range type is acquisition, a range intersection (rather than containment) check will be performed. Although the range types could be repeated, there is no benefit to repeating the same range type.

All of the restriction fields (dataset ID, ranges, online flag, browse flag, and visibility) will be joined together with the Boolean AND when the search is performed. The standard FTP URL format is:

```
ftp://[user ID:password@]host_name[:port]/[path name/][file name]
```

The file name is ignored and a unique name will be generated by ECHO to ensure uniqueness of the file name. ECHO will default to send an email notification to the requesting user when the process has completed. Data Partners may request that ECHO suppress these emails. In this case, the target ftp area will need to be monitored in order to determine when file generation has completed.